Research Article　　　　　　　　　　　　　　　**Open Access**

Microarray-based Functional Nanoproteomics for an Industrial Approach to Cancer: I Bioinformatics and miRNAome

Claudio Nicolini[1,2*], Nicola Luigi Bragazzi[1,2] and Eugenia Pechkova[1,2]

[1]Nanoworld Institute Fondazione EL.B.A. Nicolini (FEN), Bergamo, Italy

[2]Laboratories of Biophysics and Nanobiotechnology (LBN), Department of Experimental Medicine (DIMES), University of Genoa, Via Pastore 3, 16132, Genova, Italy

*Correspondence to:

Professor Claudio Nicolini
President, Nanoworld Institute Fondazione
EL.B.A. Nicolini (FEN), Largo Redaelli 7
Pradalunga, Bergamo 24020, Italy
Tel/Fax: +39 035767215
E-mail: president@fondazioneelba-nicolini.org

Abstract

Oral Lichen Planus (OLP) is a common chronic inflammatory disease, which involves the mucous membranes of the oral cavity, with an overall age-standardized prevalence of 1.27% (0.96% in men and 1.57% in women) as reported in the literature by McCartan and Healy. In our previous work, we exploited a bioinformatics approach, namely the Leader Gene Algorithm (LGA), enabling to underpin the main hub genes (termed as Leader Genes) involved in biological processes. In the case of OLP, we found a complex network made up of 132 genes and, in particular, we found five Leader Genes (namely, JUN, EGFR, FOS, IL2, and ITGB4). Using a subsequent bioinformatics algorithm, we managed to find the 48.39% of the already established OLP-related microRNAs (miRNAs), suggesting that at least half of the OLP-related microRNAome (miRNAome) finely tunes few, highly interconnected hub genes. Now, we would use real clinical samples in order to validate our predicted biomarkers, using molecular biology techniques, mass-spectrometry (MS) and *ad hoc* in-house developed instruments, such as Nanoconductimetry via Quartz Crystal Microbalance with Dissipation factor monitoring (QCM_D). A unique combination of genomics and proteomics approaches can indeed represent a promising innovation for a personalized treatment of OLP and oral cancer.

Keywords

Introduction

Oral Lichen Planus (OLP) is a common chronic inflammatory disease [1], which involves the mucous membranes of the oral cavity, with an overall age-standardized prevalence of 1.27% (0.96% in men and 1.57% in women), as reported in the literature by McCartan and Healy [2].

The histopathological features of OLP are hydropic degeneration of the basal cell layer, hyperkeratosis, acanthosis, irregular ridges and a dense band-like infiltration of T lymphocytes mainly in the *lamina propria*. Although the aetiology of OLP is still unknown, it has been widely accepted that immunological impairments are very critical among the multiple aetiological factors. Previous studies have suggested that it may represent a cell-mediated immunologic response to an induced antigenic change in the oral mucosa [3, 4].

In our previous work [5], we exploited a bioinformatics approach, namely the Leader Gene Algorithm (LGA), enabling to underpin the main hub genes (termed as Leader Genes) involved in biological processes. Only human genes were considered. In this way, it was possible to identify a list of candidate genes potentially involved in OLP pathogenesis, by mining different genomics databases and datasets.

The preliminary set of genes was then expanded using the web-available software STRING (version 9.1, freely available at http://string-db.org/) [6], considering only direct interactions (i.e. physical contact between encoded proteins, gene expression microarray data, or direct linkage in the same pathway), with a high degree of confidence (above 0.9 – confidence value in STRING ranges between 0 and 0.99, with 0.99 being the highest confidence). In this way, it is possible to identify new genes directly linked to those with an already established role in OLP, and therefore potentially involved in this disease. In order to discard false positives, results were then filtered using a further search in literature and gene databases. The process was repeated until no new gene potentially involved in OLP was identified. Then, an interaction map among the identified genes was calculated using STRING. This software can give a combined association score to each interaction, representing the degree of confidence for each interaction. For every gene identified, we summed the different combined association scores with the other genes. The sum of all these scores is defined as the weighted number of links (WNL).

Genes were then clustered, using hierarchical or K-means algorithms, according to their WNL. The genes belonging to the highest rank are defined as Leader Genes; these genes have a significant higher WNL if compared with the other ones. The other ranks are termed class B, class C, class D genes and so on, according to their WNL scores. Genes with no identified interactions (i.e. WNL=0) are defined as orphan genes. Differences among various classes in terms of WNL were statistically evaluated using an Analysis of Variance (ANOVA) test, with a Tukey-Kramer test as *post-hoc* test. Statistical significance was set at a p-value < 0.001, in order to ensure a high level of data reliability.

LGA was already successfully applied for investigating hub genes in different pathologies, such as periodontitis [7], dental diseases and their risk of malignant transformation [8, 9], and kidney transplantation [10, 11], among others.

Moreover, interacting genes were classified as up-regulated, down-regulated or neutral in respect to OLP pathogenesis. For neutral genes, we mean that they do not exhibit fold expression changes in the disease *versus* health control condition or genes for which there is not a universal consensus in the literature and in the databases.

In the case of OLP, we found a complex network made up of 132 genes and, in particular, we found five Leader Genes (namely, JUN, EGFR, FOS, IL2, and ITGB4). The interactions in the obtained network showed power law behavior, in good agreement with the scale-free topology theory of the biological graphs.

Interestingly, all of them but EGFR were up-regulated.

Evidence concerning EGFR regulation is indeed controversial and not conclusive. Moreover, the Leader Genes were widely distributed in the network (in term of topological parameters, such as stress, eccentricity and radiality) and showed higher topological coefficients than the other genes.

Bioinformatics Section

In the current work, in order to predict the potential microRNAs (miRNAs) network related to OLP, we used our previously identified "Leader Genes", namely both the Class A genes (JUN, EGFR, FOS, IL2, and ITGB4) and the Class B genes (CASP3, CD247, IL2RA, IFNG, MMP2 and LAMC2).

Genetic and genomics research is rapidly increasing our understanding of the molecular basis of some diseases and may also suggest new diagnostic and treatment strategies, paving the way for technological and industrial innovations. Many oral diseases are complex, multifactorial disorders and, as such, have a genetic basis. Studies of these pathological conditions suggest that multiple gene interactions are important determinants of susceptibility.

miRNAs are a family of small and short (usually 19-25 nucleotides long), single-stranded, endogenous, non-coding RNA molecules. They play a key role both in physiology and in pathology, and their role in pathogenesis of oral diseases is emerging. However, despite the importance of incorporating the oral microRNAome (miRNAome) in the study of oral disorders, so far only few miRNAs related to them have been discovered and described in the literature. For this reason, we predicted OLP-related miRNAs.

This could accelerate research in the field, stimulating researchers to test and verify our hypothesis and comparing wet-lab findings with computational predictions.

Then, we mined the miRGen database using the "Targets" option (freely available at http://www.microrna.gr/mirgen). This software relies on a relational database technology and enables a comprehensive research, integrating both the most widely used databases and repositories with the list of experimentally validated miRNAs and the bioinformatics tools for the miRNAs prediction (namely, DIANA-microT, miRanda, miRBase, PicTar, and TargetScanS) [12, and references therein]. We checked the biological significance of our obtained miRNAs networks mining the extant literature and using *ad hoc* bioinformatics tools (such as the miR2Disease Base, freely accessible at http://watson.compbio.iupui.edu:8080/miR2Disease/searchDiseasePre.jsp) [12, and references therein].

In order to verify the statistical significance of the enrichment of our miRNA-related list, we randomly generated a list of 11 genes (5 for Class 1 and 6 for Class 2). We used the RSA tool for this purpose (freely accessible at http://rsat.ulb.ac.be/rsat/random-genes_form.cgi), selecting "*Homo sapiens*" as source organism.

Using the above-mentioned tools we generated a network of miRNAs associated to the obtained list of random genes.

The two miRNAomes were compared using the statistical test for comparing two proportions/percentages. This computation was done with the commercial software MedCalc (version 15, down-loadable at http://www.medcalc.org/) and using the R environment (freely available at http://www.r-project.org/).

Topological properties of the obtained graphs portraying the OLP-related miRNAomes have been also studied. We investigated the clustering coefficient (a measure of degree to which nodes in a graph tend to cluster together), the network diameter and radius, the network centralization, the number of shortest paths (in percentage), the characteristic path length, the average number of neighbors, and the network density (the proportion of all the possible ties that are actually present). Network density reflects the speed at which signaling information diffuse among the nodes.

Table 1: Comparison of the topological properties among the entire OLP-related miRNAome, and the ones associated to Class A and B leader genes.

Topological parameters	Entire OLP-related miRNAome	Class A Leader Genes-related miRNAome	Class B Leader Genes-related miRNAome
Clustering coefficient	0.19	0.96	0.73
Network radius	3	2	2
Network diameter	2	1	1
Network centralization	0.94	0.99	0.99
Shortest paths (%)	99	99	99
Characteristic path length	1.99	1.99	1.99
Average number of neighbors	41.49	5.78	7.70
Network density	0.04	0.01	0.01

All the parameters were studied using Cytoscape software (freely downloadable at http://sourceforge.net/projects/graph-medusa/) [12, and references therein]. Networks are visualized using Medusa software [12, and references therein].

Results

miRNAs that have been already experimentally validated represent the 35.48% (11/31) of the entire list. Class A leader genes-related miRNAs predict the 54.17% (104/192) of the OLP + oral diseases-associated miRNAs validated in the literature. Compared to the related randomly generated miRNAome, the numbers of experimentally validated miRNAs did not statistically differ (9/31 *versus* 11/31), but the numbers of oral disease-associated miRNAs yielded a statistical significance (p-value < 0.05).

miRNAs that have been already found to be associated with OLP are the 32.26% (12/31) of the entire list, against 9/31 for the randomly generated miRNAs (not statistically significant). Class B leader genes-related miRNAs predict the 46.03% (58/126) of the OLP + oral diseases-associated miRNAs validated in the literature. Taken together the Class A and Class B leader genes-related miRNAs, they can

predict the 48.39% (15/31) of the OLP and the of OLP + oral diseases-associated miRNAs validated in the literature.
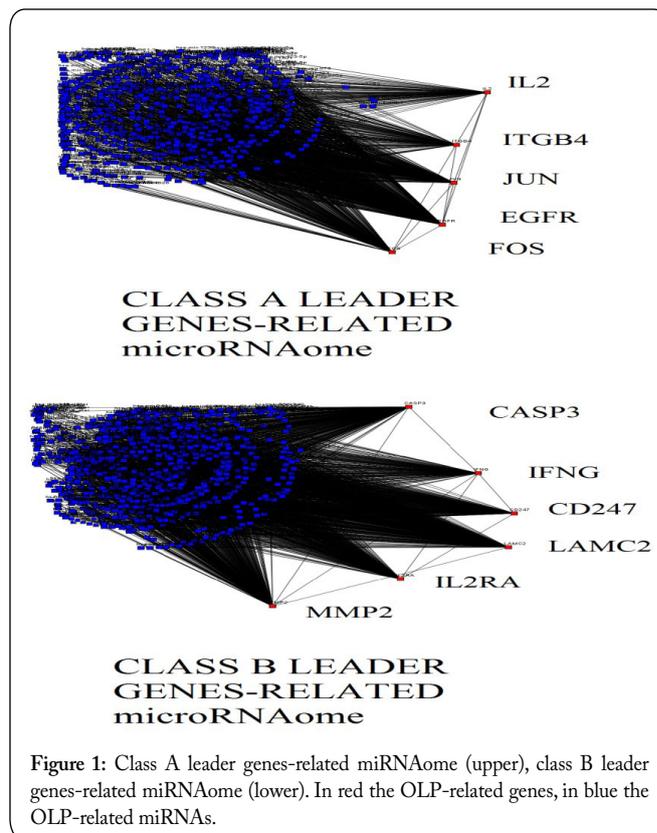

Figure 1: Class A leader genes-related miRNAome (upper), class B leader genes-related miRNAome (lower). In red the OLP-related genes, in blue the OLP-related miRNAs.

In Table 1, the topological properties of the OLP-related miRNAome are reported. The two sub-sets related to Class A and Class B Leader Genes are shown in Figure 1.

Here, in the current contribution, we introduced a novel biomolecular strategy for predicting diseases-related miRNAs, based upon our previous published LGA. We managed to find most of the already established OLP-related miRNAs. Moreover, we found also other miRNAs that have not been directly linked with OLP yet, but have a role in other oral diseases. For the remaining miRNAs their role is still unknown.

This biomolecular strategy can foster further research in the field of the biology of oral diseases, suggesting to the researchers and molecular biologists some targets to focus on and to explore with *ad hoc* targeted experiments. Moreover, it may lead to new bioinformatics software for both commercial and academic applications.

Conclusions

These data could further confirm that an approach based on bioinformatics and data-mining of existing databases could be a starting point to improve our knowledge about cellular processes and molecular mechanism of diseases and to plan targeted experimentation. In particular, the detailed analysis of gene interaction maps and the ranking of genes according to their number and confidence of interactions might have great value in the identification of new targets for a focused experimental analysis, which may confirm each hypothesis and

suggest potential risk factors and therapy targets. Noteworthy, a proper combination of experimental and theoretical results is necessary to draw a significant picture of a complex phenomenon, such as gene expression in a particular biologic system. In this study, some genes with a potential major role in OLP were identified and are preliminarily divided into three different groups according to their function. Their miRNAs are predicted and studied. These results might suggest targeted DNA or protein micro-arrays such as SNAP micro-arrays, as well as RT-PCR and mass spectrometry (MS) experiments and Nanoconductimetry via Quartz Crystal Microbalance with Dissipation factor monitoring (QCM_D) [13-15], focused on significant genes and simpler to be analyzed than mass scale molecular genomics.

There are many relevant technological and industrial implications of our study: from the possibility of creating panels of salivary or serum biomarkers [16], or of using stem cell therapy delivering particular genes for healing oral mucosal lesions [17]. A unique combination of genomics and proteomics approaches can indeed represent a promising innovation for a unique personalized treatment of OLP and oral cancer [18].

This theoretical premise is unique for its potential clinical value, considering that the marginal (although significant) differences between randomly generated and experimentally validated miRNAs point to the critical importance of prospective validation in a patient cohort [19]. Assuming the prevalence rate reported in reference [2] in the range 1.27-1.57% and a precision of 5%, a minimum number of 20-24 patients, within the 95% confidence interval, is needed to validate our bioinformatics predictions. This will be the focus of a separated, ongoing, clinical assessment.

Acknowledgments

References

1. Anuradha CH, Reddy BV, Nandan SR, Kumar SR. 2008. Oral lichen planus. A review. *N Y State Dent J* 74(4): 66-68.

2. McCartan BE, Healy CM. 2008. The reported prevalence of oral lichen planus: a review and critique. *J Oral Pathol Med* 37(8): 447-453. doi: 10.1111/j.1600-0714.2008.00662.x

3. Sugerman PB, Savage NW. 2002. Oral lichen planus: causes, diagnosis and management. *Aust Dent J* 47(4): 290-297. doi: 10.1111/j.1834-7819.2002.tb00540.x

4. Sugerman PB, Savage NW, Walsh LJ, Zhao ZZ, Zhou XJ, et al. 2002. The pathogenesis of oral lichen planus. *Crit Rev Oral Biol Med* 13(4): 350-365. doi: 10.1177/154411130201300405

5. Orlando B, Bragazzi N, Nicolini C. 2013. Bioinformatics and systems biology analysis of genes network involved in OLP (Oral Lichen Planus) pathogenesis. *Arch Oral Biol* 58(6): 664-773. doi: 10.1016/j.archoralbio.2012.12.002

6. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, et al. 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 41(Database issue): D808-815. doi: 10.1093/nar/gks1094

7. Covani U, Marconcini S, Giacomelli L, Sivozhelevov V, Barone A, et al. 2008. Bioinformatic prediction of leader genes in human periodontitis. *J Periodontol* 79(10): 1974-1983. doi: 10.1093/nar/gks1094

8. Giacomelli L, Oluwadara O, Chiappe G, Barone A, Chiappelli F, et al. 2009. Relationship between human oral lichen planus and oral squamous cell carcinoma at a genomic level: a datamining study. *Bioinformation* 4(6): 258-262.

9. Giacomelli L, Covani U. 2010. Bioinformatics and data mining studies in oral genomics and proteomics: new trends and challenges. *Open Dent J* 4: 67-71. doi: 10.2174/1874210601004020067

10. Bragazzi NL, Nicolini C. 2013. A leader genes approach-based tool for molecular genomics: from gene-ranking to gene-network systems biology and biotargets predictions. *J Comput Sci Syst Biol* 6: 165-176. doi: 10.4172/jcsb.1000113

11. Racapé M, Bragazzi N, Sivozhelezov V, Danger R, Pechkova E, et al. 2012. SMILE silencing and PMA activation gene networks in HeLa cells: comparison with kidney transplantation gene networks. *J Cell Biochem* 113(6): 1820-1832. doi: 10.1002/jcb.24013

12. Bragazzi NL, Pechkova E, Nicolini C. 2014. Proteomics and proteogenomics approaches for oral diseases. *Adv Protein Chem Struct Biol* 95:125-162. doi: 10.1016/B978-0-12-800453-1.00004-X

13. Nicolini C, Bragazzi NL, Pechkova E. 2015. Determination of protein-protein interaction for cancer control via mass spectrometry and nanoconductimetry of NAPPA SNAP arrays: an overview. *NanoWorld J* 1(1): 9-17. doi: 10.17756/nwj.2015-002

14. Spera R, Festa F, Bragazzi NL, Pechkova E, LaBaer J, et al. 2013. Conductometric monitoring of protein-protein interactions. *J Proteome Res* 12(12): 5535-4557. doi: 10.1021/pr400445v

15. Nicolini C, Adami M, Sartore M, Bragazzi NL, Bavastrello V, et al. 2012. Prototypes of newly conceived inorganic and biological sensors for health and environmental applications. *Sensors (Basel)* 12(12): 17112-1727. doi: 10.3390/s121217112

16. Totan A, Miricescu D, Parlatescu I, Mohora M, Greabu M. 2015. Possible salivary and serum biomarkers for oral lichen planus. *Biotech Histochem* 90(7): 552-558. doi: 10.3109/10520295.2015.1016115

17. Suma GN, Arora MP, Lakhanpal M. 2015. Stem cell therapy: a novel treatment approach for oral mucosal lesions. *J Pharm Bioallied Sci* 7(1): 2-8. doi: 10.4103/0975-7406.149809

18. Nicolini C, Bragazzi N, Pechkova E. 2012. Nanoproteomics enabling personalized nanomedicine. *Adv Drug Deliv Rev* 64(13): 1522-1531. doi: 10.1016/j.addr.2012.06.015

19. Bragazzi NL, Nicolini C. 2015. Lichen Planus. In: Sonis ST (eds) Genomics, Personalized Medicine and Oral Disease. Springer, USA, pp 185-217.